Method for selecting packets in a data transmission network

This invention relates to the transmission of information in a data packet transmission network of the TCP/IP type and more particularly the handling of TCP/IP packets.

It is well known that the transmission of information on the Internet using the TCP/IP protocol takes place in the form of a transmission of packets or datagram of specified standard format. Connection between a transmitting equipment and a receiving equipment is therefore characterised by a succession of packet exchanges whose function is first to initialise the connection and then transmit the data exchanged.

However these packets are routed in a variable way depending upon the load on the network. As a consequence packets which have been transmitted before others may be received after them.

In order to permit the receiving equipment to know the order in which the packets were transmitted and therefore the logical order of the data which they contain, the TCP/IP protocols provide that each TCP/IP packet contains a sequence number or count value in its header block.

As a general rule this value is initialised by the transmitting equipment at the start of the connection a random value for security purposes so that third parties cannot without difficulty break into the exchange simulating the creation of "authentic" packets. This number is then incremented by 1 or the number of bits of data transmitted in the packet whenever a new packet is transmitted.

In addition to this the problem of optimising the network's resources to the load upon it and therefore increasing the performance of the equipment processing the packets such as, for example, routers, regularly arises.

Various investigations have demonstrated that by allowing small items of work to be processed with priority over long items of work in a system with a limited resource, such as an IT system, a router or a data transmission link, the average performance of the system can be greatly improved in comparison with other forms of handling.

The question arising is therefore to distinguish between small items of work and long items. Now at the present time this has proved to be virtually impossible while processing is in progress.

One relatively effective approximation therefore consists of giving priority treatment to items of work which have received the least attention, for example, which have taken up the least calculation time.

In the case of data transmissions this is tantamount to considering that connections which have been processed for the longest time without being terminated are on average longer than others.

Investigations have in fact shown that in the case of the Internet, one effective optimisation parameter has been the volume of information transmitted, or its equivalent, the number of packets transmitted; a file whose transmission is not always completed after the transmission of a volume X

will on average be longer than a file which is not always terminated after the transmission of a volume x, where x is smaller than X.

Various mechanisms have been put forward to make use of this property in order to improve the average performance of a system. Application of a priority handling mechanism is based on handling of the volumes already transmitted by the connections. This is carried out through the use of counters, one per connection, which store the volumes of data already transmitted in memory and make a comparison between them or with a predetermined value in order to allocate processing priorities.

These systems have the disadvantage that they need storage capacity so that one counter can be installed per connection, and powerful information processing facilities because when each packet arrives the corresponding counter has to be updated, or created in the case of a new connection, compared to determine the priority which has to be allocated to that packet and, when the connection is completed (an event which it is often difficult to determine), the counter has to be closed down.

The object of the invention is therefore to overcome these disadvantages by offering a mechanism which is simple and quick to implement.

The subject matter of the invention is therefore a method for the selection by a downstream device of data packets of connections of a network transmitted by at least one upstream device in relation to a predetermined threshold of

the quantities of data transmitted by these connections, this method comprising
- at the start of each connection, initialisation of a transmitted packets counter by each upstream device to an initial count value,
- incrementing the said counter by a specified value for each packet transmitted, defining the current count value of the packet, and copying this current count value into the packet header block by the upstream device,
- reception by the downstream device of each IP packet of each connection, characterised in that the method comprises:
- selecting the initial count value at the upstream device from a set of predetermined initial values such that the difference between two consecutive initial values in that set is greater than the predetermined threshold, and
- comparing the predetermined threshold in the downstream device with the difference between the current count value and the immediately lower initial value in the set of predetermined initial values, as a result of which the packets corresponding to the connections which have transmitted less data than the predetermined threshold can be selected in preference to the packets corresponding to connections which have transmitted more data than the predetermined threshold.

According to the invention, when the current count value is presented in the binary form of a recording of n bits, the set of initial count values is such that a field of l bits of the count value, l always being less than n, is systematically initialised to zero, this field being positioned in such a way that when the number of transmitted packets reaches the predetermined threshold at least one bit in this field takes the value 1.

According to the invention, if the field of l bits is positioned between a bit of rank m and a bit of rank m + l in the count value, the initial count values will be greater than 2^{l+m}.

According to the invention, the predetermined threshold is equal to 2^m-1.

According to the invention, the initial count values are multiples of 2^{l+m}.

According to the invention, the bits of low weight in the initial count values are selected randomly from the bits of rank below t, t always being less than m.

According to the invention, the number of bits of low weight t is the whole part of the base 2 logarithm of the maximum packet size permitted on the network.

According to the invention, the comparison by the downstream equipment is a comparison between the field of l bits and 0.

According to the invention, for the packets corresponding to connections having quantities of transmitted data which are less than the predetermined threshold, it comprises allocating them a processing priority which is higher than the packets corresponding to connections having quantities of transmitted data which are greater than the predetermined threshold.

The invention further relates to a system for generating connection of a network data packets, in an upstream device

connected to the network, comprising means for the transmission of packets connected to the network, these transmission means being connected to an information processing unit and information storage means comprising at least one register capable of storing the number of transmitted packets, the information processing unit comprising means for initialising this register to an initial count value at the start of the connection and means for incrementing the contents of this register whenever a new packet is created and means for copying this register into a current count value field in the packet header block, characterised in that the initialisation means comprise means for selecting the initial count value of each connection from a set of predetermined initial values such that the difference between two consecutive initial values in that set is greater than a predetermined threshold.

According to the invention, if the current count value is in the binary form of a recording of n bits, the set of initial count values is such that a field of l bits of the count value, where l is always smaller than n, is systematically initialised to 0.

According to the invention, the incrementing means comprise means for setting at least one bit in the field of l bits to the value of 1 when the number of packets transmitted exceeds the predetermined threshold.

According to the invention, if the field of l bits is positioned between a bit of rank m and a bit of rank m+l of the count value, the initial count values are greater than 2^{l+m}.

According to the invention, the predetermined threshold is equal to 2^{m-1}.

According to the invention, the initial count values are multiples of 2^{l+m}.

According to the invention, the means for selecting the initial count values comprise means for random selection of the low weight bits of the initial count values from the bits of rank lower than t, t always being smaller than m.

According to the invention, the number of low weight bits t is the whole part of the base 2 logarithm of the maximum packet size permitted on the network.

The invention further relates to a system for the transmission of data packets from at least one connection of a network comprising receiving means for packets originating from upstream device and packet transmission means connected to information processing means, each packet having a current count value in its header block, characterised in that the information processing means also comprise a table of initial count values and means for calculating the difference between the current count value of the packet received and the initial value in the table immediately below that current count value and means for comparing this difference with a predetermined threshold.

According to the invention, the means for calculating the difference and comparison with the predetermined threshold make a comparison between a field of l bits of the current count value and zero.

According to the invention, the information processing and transmission means give priority in processing to the packets corresponding to connections having quantities of transmitted data which are less than the predetermined threshold over packets corresponding to connections having quantities of transmitted data greater than the predetermined threshold.

The invention will be better understood from a reading of the following description provided by way of example with reference to the appended drawings in which:
- Figure 1 shows an overall diagram of a TCP/IP network,
- Figure 2 shows a simplified diagram of a TCP packet header block according to the standard,
- Figure 3 shows a symbolic diagram of the distribution of initial values according to the invention,
- Figure 4 shows a diagram of the count value according to the invention in a binary representation,
- Figure 5 shows a diagram of an upstream packet generation system,
- Figure 6 shows a diagram of a downstream packet transmission system.

In a TCP/IP network like the Internet, Figure 1, upstream computers 1 establish connections with computers 2 through the intermediary of network equipment such as routers 3a, 3b, 3c whose function is to route the packets constituting the connection between computers 1 and 2. Network equipment 3a, 3b, 3c will be referred to as downstream equipment in the remainder of the description.

A packet or datagram, Figure 2, comprises a header block 5 comprising the predetermined fields which are necessary for

satisfactory routing of this packet such as, for example, the address of the transmitting computer and that of the receiving computer, and the data 6 transmitted.

Among the many fields in header block 5 there is a counting field 7. This is generated by the transmitting equipment from an initial count value by incrementing this by either one unit per packet transmitted or the number of bits of data transmitted since the start of the connection.

Receiving equipment 2 can therefore check that it has correctly received all the data and put them back into order if necessary.

In the implementation of the method according to the invention described here, the initial count value is selected randomly from a restricted predefined number of count values PIN_1, PIN_2, ..., PIN_n, thus constituting a set of predetermined initial values.

As explained below, the difference between two consecutive initial values must be very much greater than the predetermined threshold value th which is used to distinguish the packets.

This list of initial count values is also known by the downstream equipment. Thus when a packet arrives the downstream equipment estimates the volume transmitted by the connection by taking the difference between the count value for the packet and the initial count value in the list immediately beneath it.

If the volume so estimated is less than the predetermined threshold value th the packet forms part of a short connection and therefore receives high priority, otherwise it forms part of a "long" connection and then receives a low priority.

It is worthwhile noting that if the connection is sufficiently long the current count value can reach and exceed the initial count value PIN_{i+1} which is immediately above the initial count value PIN_i in the list used to start the connection. Because of this the downstream equipment then calculates the difference between the current count value and PIN_{i+1} and not PIN_i, which has the effect of incorrectly classifying the packet as belonging to a short connection, and this will continue as long as this new difference is not greater than the predetermined value th.

This incorrect classification has in fact little impact on the function of priority management. In fact the distribution of connection volumes follows a Pareto law. Thus 80% of the volume is generated by 20% of connections. By judiciously setting the predetermined threshold, these 20% or less will be regarded as long connections.

On the other hand the error does not occur when a number of packets equal to or greater than $PIN_{i+1} - PIN_i$ have already been transmitted, a difference which is very much greater than the threshold value predetermined by judicious choice of these initial count values.

As a consequence packets which are incorrectly classified are doubly rare: they belong to rare connections and they

represent a small proportion of the packets transmitted by these connections.

The invention thus advantageously makes it possible to sort packets in relation to the volume transmitted by the corresponding connection without having to operate a counter specific to that connection in each downstream equipment.

A variant implementation of the invention comprises selecting the initial count values so that a zone of l bits of the counter is always initialised to zero.

In order to do this, Figure 4, the counter and its associated values may be represented in the form of a field of n bits and may therefore take 2^n separate values.

Furthermore, the predetermined threshold value th is selected to be equal to a power of 2 minus 1, that is 2^m-1. In binary representation th has the bit in the m position set to 0 and the other bits of lighter weight set to unity.

As an assumption, the predetermined threshold value th is very much lower than the maximum possible value for the counter, that is 2^n.

As explained above, the initial count values must also be very much greater than th. As an assumption it is therefore possible to select them to be greater than 2^{l+m} and such that the field between the bit of weight m and the bit of weight l+m-1 is initialised to zero. These are selected preferentially from the multiples of 2^{l+m}.

Thus in the course of the method of incrementing the count value at least one of the field bits [2^m, 2^{l+m}] is set to 1 when the predetermined threshold value is reached.

Thus the difference operation performed by the downstream equipment takes the form of a comparison with zero in the corresponding field to determine whether the packet belongs to a short or a long connection, without it being necessary to maintain a table of initial count values.

Priority management is therefore effected through an extremely simple calculation which very advantageously resolves the problems arising in the prior art.

In the embodiment of the invention described above, the low weight bits of the initial count values, that is to say those having a weight of less than m are also initialised to zero. There are therefore $2^{n-(l+m)}$ possible initial values because they have been selected from multiples of 2^{l+m}.

However for security reasons it may be necessary to increase the number of possible initial count values in order to reduce the probability of selecting two identical numbers. In fact if the counter is a packet counter, it is possible to chose th and therefore m to be small, and therefore the number of initial values remain sufficiently high for random selection to be effective. On the other hand, if the counter is a data bit counter as in the TCP standard, th must be very much larger and this correspondingly restricts the number of initial count values.

By way of example, for packets of 1024 bits, which is a value very frequently found on the Internet, a connection is

regarded as being long when it comprises more than 5 packets. For packet counting this means that m is selected to be equal to 2 or 3. But when transmitted data bits are counted m becomes 12 or 13.

In this latter case a variant embodiment of the invention comprises selecting a certain number of low weight bits in the initial sequence number. By choosing weights which are very much less than that of the predetermined threshold value, for example from the bits of lesser weight than the whole part of the base 2 logarithm of the maximum size for a data packet permitted on the network, only a minimum error in exceeding the predetermined threshold value for the corresponding connection is introduced. In fact, considering again the example of a data bit counter functioning for packets having more than 1024 data bits, if the bits of weight less than 10 are selected randomly, it is easy to see that the error introduced will not apply to more than one packet.

Thus the method according to the invention advantageously makes it possible to simplify the detection of connection sizes while maintaining random selection of the initial values.

In order to implement the method according to the invention the upstream equipment transmitting the packets, Figure 5, comprises transmission means 8 for these packets connected to the network, such as, for example, an Ethernet network card. An information processing unit 9 is connected to these transmission means 8 and information storage means 10, such as for example volatile memory. An address in this memory acts as register 11 for counting the packets or bits

transmitted. Information processing unit 9 comprises means 9a for initialising this register 11 at the start of the connection to an initial count value selected in the manner explained above. These initialisation means 9a also comprise means 9d for selecting the initial value from a set of predetermined values and means 9e for randomly selecting the low weight bits of this initial value. Information processing unit 9 also comprises means 9b for incrementing the contents of this register 11 whenever a new packet is transmitted, this increment being either one unit in the case of packet counting or the number of data bits transmitted in the packet. In the preferred embodiment of the invention these incrementing means 9b comprise means 9f for setting at least one bit in the field of l bits to the value 1. Information processing means 9 also comprises means 9c for copying the value of this register 11 to the field for the current count value in the packet header block.

The downstream equipment, Figure 6, is a packet transmission system. It therefore comprises means 12 for receiving packets originating from the upstream equipment and means 13 for transmitting these packets to their final destination. These reception means 12 and transmission means 13 are connected to an information processing unit 14 which also comprises a table 15 of initial count values and means 14a for calculating the difference between the current count value of the packet and the initial value in the table immediately below it. This information processing unit 14 also comprises means 14b for comparing this difference with the predetermined threshold values th.

Of course those skilled in the art will be in a position to apply the invention to any network protocol without difficulty using packets numbered in transmission order.